Multi Locus Sequence Analysis of root nodule isolates from *Lotus arabicus* (Senegal), *Lotus creticus*, *Argyrolobium uniflorum* and *Medicago sativa* (Tunisia) and description of *Ensifer numidicus* sp. nov. and *Ensifer garamanticus* sp. nov.

C. Merabet, M. Martens, M. Mahdhi, F. Zakhia, A. Sy, C. Le Roux, O. Domergue, R. Coopman, A. Bekki, M. Mars, A.Willems and P. de Lajudie.

In: International Journal of Systematic and Evolutionary Microbiology 60 (3), 664-674, 2010.

http://ijs.sgmjournals.org/cgi/content/full/60/3/664

Merabet C., M. Martens, M. Mahdhi, F. Zakhia, A. Sy, C. Le Roux, O. Domergue, R. Coopman, A. Bekki, M. Mars, A.Willems and P. de Lajudie. 2010. Int. J. Syst. Evol. Microbiol. 60:664-674. DOI 10.1099/ijs.0.012088-0

IJS/2009/012088 REVISED MS

Multi Locus Sequence Analysis of root nodule isolates from *Lotus arabicus* (Senegal), *Lotus creticus*, *Argyrolobium uniflorum* and *Medicago sativa* (Tunisia) and description of *Ensifer numidicus* sp. nov. and *Ensifer garamanticus* sp. nov.

C. Merabet[1,2], M. Martens[3], M. Mahdhi[2,4], F. Zakhia[2], A. Sy[5], C. Le Roux[2], O. Domergue[2], R. Coopman[3], A. Bekki[1], M. Mars[4], A.Willems[3] and P. de Lajudie[2*].

[1]Laboratoire de Rhizobiologie, Univ. Oran, Es-Senia, Algeria.

[2]IRD, UMR 113 Symbioses Tropicales et Méditerranéennes F-34398 Montpellier, France.

[3]Laboratorium voor Microbiologie (WE10), Universiteit Gent, Belgium.

[4]Faculté des Sciences, Gabès Tunisia.

[5] IRD-UCAD-ISRA, Laboratoire Commun de Microbiologie, Dakar, Senegal, West Africa.

***Corresponding author:** Philippe.Lajudie@ird.fr

L.S.T.M., UMR 113 IRD/CIRAD/AGRO-M/UM-II, Campus International de Baillarguet, TA A-82/J, 34398 MONTPELLIER Cedex 5, France.
Tel. : (33) (0) 4 67 59 38 51
Fax : (33) (0) 4 67 59 38 02

Running title: *Ensifer numidicus* and *Ensifer garamanticus* spp. nov.

Subject category: New Taxa (subsection Proteobacteria).

The GenBank/EMBL/DDBJ accession numbers of new sequences are provided in Table 1.

Abstract:

Nine isolates from *Argyrolobium uniflorum*, *Lotus creticus*, *Medicago sativa* (Tunisia) and *Lotus arabicus* (Senegal) were analysed by Multilocus Sequence Analysis (MLSA) of five housekeeping genes (*recA, atpD, glnA, gltA* and *thrC*), the 16S rRNA gene and the nodulation gene *nodA*. Analysis of the individual gene sequences and concatenated gene sequences demonstrated that the nine strains constitute three stable, well-supported (bootstrap and sequence similarity values) monophyletic clusters A, B and C, all belonging to the *Ensifer* genus branch, whatever the phylogenetic reconstruction method (Maximum likelyhood, Maximum Parsimony, Neighbour Joining). The three groups were further characterized by API 100 auxanographic tests, host specificity and *nodA* gene sequence analysis. In light of all data we describe clusters A and C as two novel species within the genus *Ensifer*, *Ensifer numidicus* sp. nov., with ORS 1407^T (=LMG 24690^T =CIP 109850^T) as the type strain, and *Ensifer garamanticus* sp. nov., with ORS 1400^T (=LMG 24692^T =CIP 109916^T) as the type strain. Cluster B strains are assigned to *Ensifer adhaerens* genomovar A.

Keywords: rhizobia, *Ensifer numidicus*, *Ensifer garamanticus*, MLSA.

INTRODUCTION

During the past three decades, bacterial taxonomy was based on a polyphasic approach including both phenotypic and genotypic data with 16S rRNA gene sequence analysis and DNA-DNA hybridisation being considered as the cornerstones of genotypic characterisation for new species description (Vandamme *et al*., 1996; Stackebrandt *et al*., 2002; Coenye *et al*., 2005). However many authors have reported drawbacks of the methods (Stackebrandt *et al*., 2002; Stackebrandt, 2003; Eardly *et al*., 2005; van Berkum *et al*., 2003; Ueda *et al*., 1999; Schouls *et al*., 2003). The *ad hoc* committee for the re-evaluation of the species definition in bacteriology (Stackebrandt *et al*., 2002; Gevers *et al*., 2005) suggested the sequence analysis of a set of protein-encoding genes as alternative phylogenetic markers, so-called MLSA (Multilocus Sequence Analysis). Several recent studies confirmed that sequences from housekeeping genes can be used for identification at the species level (Zeigler, 2003; Wertz *et al*., 2003), evolutionary population genetics and in taxonomy (Bailly *et al*., 2006; Stepkowski *et al*., 2003; Vinuesa *et al*., 2005a,b). The analysis of a small number of carefully selected gene sequences may equal or even surpass the precision of DNA-DNA hybridisations for quantification of genome relatedness and this approach thus has the potential to replace the cumbersome DNA-DNA hybridisations (Zeigler, 2003, Martens *et al*., 2008). Mantelin *et al.* (2006) included MLSA in the description of four new *Phyllobacterium* species and Vinuesa *et al.* (2005c) described new *Bradyrhizobium* species using the phylogenies of three housekeeping genes *atpD, glnII, recA* combined with other classical genotypic and phenotypic analyses. *Ensifer mexicanus* and *S. chiapanecum* were described using phenotypic analysis, phylogenies of recA, gyrA, nolR, rpoB, rrs and symbiotic genes and confidence intervals of sequence identity to estimate both inter and intra-species ; all were in correlation to DNA-DNA hybridization data (Lloret et al., 2007; Rincón-Rosales et al., 2009). Also, MLSA was applied in diversity studies of *Enterococcus*, *Aeromonas* and *Ensifer* (*Sinorhizobium*) (Naser *et al*., 2005; Soler *et al*., 2004, Bailly *et al*., 2006) in which it was reported as a highly reproducible and economical method for rapid and reliable species identification.

The genera *Sinorhizobium* and *Ensifer* were recently recognized to form a single phylogenetic clade (Balkwill, 2003; Willems *et al.*, 2003), and are now united and all *Sinorhizobium* species were transfered to the genus *Ensifer*, in line with rule 38 of the Bacteriological Code (Young, 2003; Judicial Commission, 2008). *Ensifer* currently includes 11 species (Wang *et al*., 2002; Wei *et al*., 2002; Young, 2003; Toledo *et al*. 2003). Two further species, *Ensifer*

mexicanum and *Sinorhizobium chiapanecum*, have been published but not yet validated (Lloret *et al*. 2007; Rincon-Rosales et al., 2009). Martens *et al*. (2007, 2008) recently demonstrated that the discriminative power of MLSA for species identification and delineation is higher than 16S rDNA sequence analysis and DNA-DNA hybridization within the genus *Ensifer*.

Here we performed MLSA to complete the characterisation of nine root nodule strains isolated from several legumes in Tunisia (*Argyrolobium uniflorum*, *Lotus creticus*, *Medicago sativa*) and Senegal (*Lotus arabicus)*. *Argyrolobium uniflorum* (Tribe Genisteae, Family Fabaceae) is an indigenous herb legume in the Mediterranean bassin, and is a pastoral and forage legume broadly distributed in arid and semi-arid regions of Tunisia. This plant plays an important role in soil fertility maintenance, coverage and dune stability (Ferchichi, 1996). *Lotus arabicus* L. (tribe Loteae, family Fabaceae), syn. *Lotus mossamedensis* Welw. ex Baker, is a natural annual herb legume of coastal dunes in Senegal. *Medicago sativa* is the most widely cultivated species of lucerne in the world, and its symbiotic rhizobial strains have been extensively studied. Six of the isolates have previously been analyzed by ARDRA using seven enzymes, total cell protein SDS-PAGE analysis and 16S rRNA gene sequencing (Zakhia *et al*., 2004) and results suggested their separate positions in *Ensifer* (syn. *Sinorhizobium*). According to the conclusions of Martens et al. (2007, 2008) the strains were subjected to MLSA using five housekeeping genes, most discriminative among 10 tested housekeeping genes for taxonomic purposes in *Ensifer*. Genes were analysed individually and as a concatenation using only the congruent genes estimated by ILD test (incongruence-length difference (Farris *et al*., 1995). Trees were constructed using three algorithmic methods for comparison. The characterisation was completed by auxanographic tests, DNA-DNA hybridizations, host specificity and sequence analysis of the symbiotic *nodA* gene for biovar description.

MATERIALS AND METHODS:

Bacterial strains and culture media

The strains investigated are listed in Table 1. All strains were grown on yeast mannitol agar YMA (Zakhia et al., 2004) at 28 °C. Isolates were stored at –80 °C in 50 % v/v glycerol.

Host specificity

The isolates were tested for nodulation on their original host plant. The seeds from *Medicago sativa* and *Lotus arabicus* were sterilised in 3 % calcium hypochloride and scarified by immersion in 96 % Sulphuric acid for 20 min, washed five times with sterile water and placed in agar-water 1 % w/v at 24-25 °C for germination. After 3 to 4 days, the seedlings were transferred to agar slant tubes (Bertrand, 1997) for root nodulation trials (5 plants per strain). Plants were grown under continuous light (20 Wm^{-2}) at 24 °C and inoculated with 1 ml of an exponential growth phase YM bacterial suspension. Roots were observed for nodule formation during the first 4 weeks after inoculation. Infectivity of strains isolated from *Argyrolobium uniflorum* was previously reported by Zakhia (2004).

Genomic DNA isolation*:*

Bacterial DNA was prepared as described by Zakhia *et al*. (2004) or alternatively by the alkaline lysis method (Baele *et al*., 2000).

DNA amplification

The 16S rRNA gene, internal fragments of five housekeeping genes (*atpD, recA, gltA, thrC* and *glnA)* and the *nodA* gene were amplified using primers listed in Supplementary Table S1 (available in IJSEM Online).

PCR amplification of *atpD, recA, gltA, thrC* and *glnA* genes was performed as described by Martens et al. (2007) except in a total volume of 25 µl reaction mixture. PCR amplification of 16S rRNA gene was performed according to Zakhia *et al*. (2004) and *nodA* gene according to Zakhia *et al*. (2006).

Gene sequencing and phylogenetic data analysis

PCR products of different genes were purified and sequenced as previously described (Martens et al., 2007) using primers detailed in Supplementary Table S1 (available in IJSEM Online). Consensus sequences were constructed using the Autoassembler software (Applied Biosystems). Interstrain sequence similarity (S) values were determined applying pairwise sequence alignments using BioNumerics 4.6. software. The nucleotide sequences of partial *atpD, recA, gltA, glnA, thrC* and SSU were independently aligned using the Clustal X program, version 1.8 (Thompson *et al*., 1997). Neighbour-joining (NJ), Maximum-Parsimony (MP) and Maximum Likelihood (ML) trees were constructed using PAUP* version 4.0b10

(Swoford, 2002). Reference sequences of *Ensifer* species were included for comparison (accession numbers are listed in Martens *et al*., 2008). NJ analyses (Saitou & Nei, 1987) were performed using the Kimura-2 correction (Kimura, 1980) and 1000 bootstrap replications; MP analyses were performed using the heuristic search option. For ML analyses, the optimal models of nucleotide substitution were estimated using the program MODELTEST 3.7 (Posada & Crandall, 1998), using the Akaike Information Criterion (AIC) (Posada & Buckley, 2004). The MP trees were used as starting trees for the heuristic search procedure. Bootstrap analyses were performed using 1000 replications of heuristic searches for MP and 100 replications for ML. The ILD test (Farris *et al*., 1995) implemented in PAUP* and using 1250 replicates was used to assess incongruence between datasets. A phylogenetic tree was built with the concatenated sequences of the congruent housekeeping genes.

DNA-DNA Hybridisations

DNA was prepared according to a slightly modified procedure of Marmur (1961) as described previously (Willems *et al*., 2001). Hybridization were carried out using the microplate method in which unlabelled DNA, non covalently bound to the microplate, is hybridized with biotinylated probe DNA (Ezaki *et al*., 1989; Willems *et al*., 2001; Goris *et al*., 1998). Hybridizations were performed at 45 °C in 2xSSC, 50 % formamide.

Phenotypic tests

Microscopic observations, tolerance to pH, NaCl and temperature were performed as previously described (de Lajudie et al., 1994; Mahdhi et al., 2008).

API100 galleries (BioMérieux, Montalieu-Vercieu, France) were used to test the utilisation of carbon sources (carbohydrates, amino acids, organic acids) for bacterial growth. Inoculants were obtained from 36-h YMA slant cultures. After inoculation, the galleries were incubated at 30 °C, and results were determined after 1, 2, 4 and 7 days. The results of the auxanographic tests were scored as described previously (Kersters & De Ley, 1984). Jaccard similarity coefficients were calculated and an UPGMA analysis was performed using BioNumerics 4.6.

RESULTS

Nodulation tests:

All studied strains induce efficient nodulation on their hosts of isolation. This was already reported by Zakhia (2004) for ORS 1400, ORS 1401, ORS 1407, ORS 1444 (*Argyrolobium uniflorum*), for ORS 1410 (*Medicago sativa*). Here strains ORS 2154, ORS 2133 and ORS 529, isolated from *Lotus arabicus*, were tested and are effective on their plant of isolation.

16S rRNA gene sequence analysis:

Nearly full-length 16S rRNA gene sequences (1340bp) were determined for strains ORS 1410, ORS 2133, ORS 529, ORS 2154, ORS 1401 and STM 354. The 16S rRNA gene phylogenetic trees constructed using three methods (ML, MP, NJ) resulted in the same groupings. Therefore, only the resulting ML tree is shown in Fig. 1. All studied isolates were placed unambiguously within the AlphaProteobacteria and within the *Ensifer* reference species clade. They were subdivided into three clusters (clusters A, B and C) supported by high bootstrap values and displaying high intra-cluster sequence similarity.

Cluster A (Bootstrap [BT] value 97 %) includes the isolates ORS 1444, ORS 1410 and ORS1407 and formed a homogeneous group (sequence similarity values 100 %) equally distant to *E. arboris*, *E. medicae* and *E. meliloti* (99.7 % sequence similarity).

Cluster B strains (ORS 529, ORS 2154 and ORS 2133) also shared identical 16S rRNA gene sequences and formed a significant cluster. This clade grouped together with the different *E. adhaerens* genomovars (more than 99.6-99.8 % sequence similarity) and "*S. morelense*" (99.0-99.2 % sequence similarity) reference strains, although low BT values were displayed for this group (BT value 35 %).

The strains ORS 1400, ORS 1401 and STM 354 formed a homogeneous group (Cluster C, 100 % interstrain sequence similarity values) supported by a bootstrap value of 85 %, and clustered in the vicinity of *E. terangae* (48 % BT value and 98,8 % sequence homology). The type strain of *E. mexicanum* LMG 23932^T which was described recently (Lloret *at al*., 2007) is related to these taxa. It was not yet included in our analyses, but forms a distinct subclade, sharing sequence similarity values of 97.8 to 98.8 % with clusters A, B and C and 99.5 % with *E. terangae* strains.

Sequence analysis of individual genes.

The five housekeeping gene fragments examined in this study had different lengths: *glnA* (977bp), *gltA* (681bp), *recA* (550bp), *atpD* (461bp), *thrC* (636bp). Within each group A, B or C, the three strains of the group had identical sequences for all genes tested (except for ORS 1410 of group A for which we could not amplify *thrC*). Sequence analysis of the five individual housekeeping genes resulted in different tree topologies (Fig. 2). However, the same three separate clusters A, B and C were revealed with high bootstrap support for all clusters in all gene analyses, including 16S rRNA gene. These clusters occupied different positions relative to the reference species depending on the gene considered.

In the *atpD* tree topology, *A. tumefaciens* grouped with "*S. morelense*", the 3 genomovars of *E. adhaerens,* and all strains from cluster B (100 % BT) although this cluster had low BT support (BT value 32 %). A second cluster (BT value 43 %) contained all other *Ensifer* species and clusters A and C (both clusters 100 % BT value).

In the case of *gltA*, all *Ensifer* strains together with clusters A, B and C, formed a single separate but poorly supported clade (BT value 38 %). Within this clade, all strains from cluster C formed a significant subclade with *E. terangae* (100 % BT). Another well-supported subgroup (BT value 98 %) was formed by all strains from cluster B and *E. adhaerens* gv. A.

In the *glnA* analysis, all *Ensifer* strains (and clusters A, B and C) were grouped together with high BT support (75 %) and the same two significant subclusters as in the *gltA* tree topology were found (BT values 100 and 95 % for cluster C/*E. terangae* and cluster B/*E. adhaerens* gv. A respectively).

Also in the *thrC* gene tree, all *Ensifer* strains together with clusters A, B and C, formed a single, separate but less supported clade (BT value 48 %), in which the same two subclades were again distinguished (BT values 100 and 98 % for cluster C/*E. terangae* and cluster B/*E. adhaerens* gv. A respectively). In the final single gene tree, the *recA* tree, again all *Ensifer* strains and clusters A, B and C did form a separate but poorly supported clade (BT value 27 %). A well-supported subgroup (BT value 95 %) was formed by all strains from cluster B and *E. adhaerens* gv. A.

Thus, cluster A formed a separate, well-supported group (BT value of 100 %) in all single gene phylogenies with identical sequences for the three strains (100 %). Cluster A was located at different positions in comparison to the *Ensifer* reference species and no significant clusterings were apparent. The highest observed sequence similarities between cluster A strains and the reference species ranged between 86.9 % with *E. saheli* and *E. kostiensis* for the *thrC* gene and 94.3 % with reference strains *E. meliloti* and *E. medicae* for the *atpD* gene.

Cluster B grouped with *Ensifer adhaerens*, more specifically *E. adhaerens* gv. A with high BT values in all housekeeping gene analyses, except for *atpD*. Sequence similarities with *E. adhaerens* gv. A strains ranged between 96.4 % for *thrC* to 98.4 % for *recA* gene. With both other genomovars the range was lower at 90.5% (gltA) to 95.3% (glnA).

The strains ORS 1400, ORS 1401 and STM 354 formed the monophyletic Cluster C in all gene phylogenies (BT values ranging from 98 to 100 %; sequence similarity values 100 % for all genes). This cluster displays sequence similarities with its closest neighbour *E. terangae* in the range of 92,2 % (for *recA*) to 96,4 % (for *gltA*). Cluster C together with *E. terangae* was also related to the *E. mexicanum* type strain LMG 23932^T in the analysis of three housekeeping genes (*glnA, gltA* and *recA*) (not included in Fig. 2). The sequence similarities between Cluster C strains and *E. mexicanum* ranged from 92.2 % for *recA*, over 93.7 % for *gltA*, to 95.8 % for *glnA*. Between *E. terangae* and *E. mexicanum* somewhat comparable interspecies sequence similarity values amounted to 96.6 % for *recA*, 93.5 % for *gltA* and 96.2 % for *glnA*. This indicates that *E. mexicanum, E. terangae* and the isolates from Cluster C all represent distinct genomic species.

Sequence analysis of concatenated housekeeping genes.

The *atpD* gene was found not to be congruent with the other genes (p-value<0, 01) while all other housekeeping genes showed compatible phylogenetic signals (data not shown). A tree was constructed from the concatenated sequences of the four congruent genes *recA, glnA, gltA*, and *thrC* (p-value>0, 01) estimated by ILD test (Farris *et al.*, 1995). The ILD test values ranged from 0.0120 for *thrC* and *recA* to 0.6288 for *thrC* and *glnA*. In view of the ILD test results, we concatenated the aligned sequences for *recA, thrC, gltA* and *glnA* and obtained an alignment of 2704 nucleotides (comprising 1486 invariable sites, 237 variable but parsimony uninformative sites and 981 parsimony informative sites).

The combined analysis showed a significant cluster (BT value 97 %), comprising all *Ensifer* strains together with clusters A, B and C, which was subdivided into two closely related subclusters. One well-supported subcluster (BT-value 100 %) includes *"S. morelense"*, the three *E. adhaerens* genomovars and cluster B. As in most single gene analysis, cluster B formed a significant subclade together with *E. adhaerens* gv. A. The second major subcluster (BT value 99 %) comprised all other *Ensifer* strains and clusters A and C. In analogy to most single gene trees, Cluster C formed a significant sub-branch with *E. terangae* (BT value 100 %). Cluster A strains ORS 1407 and ORS 1444 also formed a reliable cluster (BT value 100 %), which was well separated from all other genomic species. Since we did not obtain a *thrC*

sequence for ORS 1410, this strain could not be included in the concatenated housekeeping gene analysis. As indicated, all mentioned clusters were supported by higher BT values in the concatenated tree than in the single gene trees and are therefore more robust.

DNA-DNA Hybridization

To determine to which *E. adhaerens* genomovar Cluster B could be designated, we performed DNA-DNA hybridisations between two strains (ORS 529 and ORS 2133) of this cluster and three *E. adhaerens* reference strains, representing the three different genomovars (Willems *et al*., 2003) (Table 2). In agreement with the gene sequence analyses, ORS 529 and ORS 2133 display high DNA-DNA hybridisation values (99 %), proving that the strains represent the same genomic species. The hybridisation values of strains ORS 529 and ORS 2133 with *E. adhaerens* gv. B and C strains are below 70 %, indicating that the two strains from Cluster B are not *E. adhaerens* gv. B nor *E. adhaerens* gv. C. However, DNA-DNA hybridisation results between cluster B strains and the *E. adhaerens* gv. A reference strain, exceed the 70 % value which confirms the assumption drawn from MLSA that cluster B strains are *E. adhaerens* gv. A strains.

***nodA* gene sequencing**

The *nodA* gene was sequenced in strains representing both the different clusters and host plants of isolation: two cluster C strains (ORS 1400/ *Argyrolobium uniflorum* and STM 354/ *Medicago sativa*), one cluster B strain ORS529 (isolated from *Lotus arabicus*, like the two other strains of the cluster) and two Cluster A strains (ORS 1444/ *Argyrolobium uniflorum* and ORS1410/*Lotus creticus)*. The *nodA* gene could not be amplified in strain ORS 1407. The analysis of *nodA* sequences by neighbour-joining method (Fig. 3), shows that strains isolated from *Argyrolobium uniflorum*, *Lotus arabicus* and *Medicago sativa* form a separate cluster. This cluster is grouped with the *nodA* of *E. meliloti* and *E. medicae* branch with high bootstrap support. In this study we report for the first time the position of *nodA* sequences of isolates from *Argyrolobium uniflorum.*

Numerical analysis of auxanographic tests

All strains were tested for use of 99 substrates as sole carbon source for growth, using API Biotype100 galleries (Supplementary Table S2). Some discriminative features could be observed.

In contrast to its closest phylogenetic neighbours *E. medicae* LMG 19920^T, *E. meliloti* LMG 6133^T *and E. arboris* LMG 14919^T, cluster A strains do not grow on xylitol and DL-glycerate. Results for other substrates were variable between members of this cluster. About 28 substrates were differently utilized by strains ORS 1444, ORS 1410 and ORS 1407 and 13 substrates were only metabolised by ORS 1407 (dulcitol, D-lyxose, 1-O-methyl-α-D-glucopyranoside, 3-O-Methyl-D-glucopyranose, D-gluconate, L-histidine, succinate, fumarate, ethanolamine, DL-β-hydroxybutyrate, L-aspartate, L-alanine, propionate).

Cluster B strains used putrescine in contrast to the three genomovars (A, B, C) of *E. adhaerens*. Seven substrates (L (+) sorbose, dulcitol, D-tagatose, citrate, 5-keto-D-gluconate, D-gluconate, glutarate) were not metabolised by Cluster B members and diversely used among the *E. adhaerens* three genomovars.

Cluster C, close to *E. terangae* in phylogenetic analysis, differed from this species in ten substrates assimilated by *E. terangae* LMG 7834^T and not by any member of Cluster C. However, malonate is metabolised by all Cluster C strains and not by *E. terangae* LMG 7834^T.

Auxanographic data were analysed numerically (data not shown). Clusters A and C formed two related clusters but separate from the other *Ensifer species* and Cluster B was included in *E. adhaerens* as observed in our genotypic analysis, with strain ORS 2154 close to *Ensifer adhaerens* gv B strain R-7457.

DISCUSSION

In this study we performed MLSA to characterise nine strains from Tunisia and Senegal isolated from root nodules of several legumes. Previously, six of them had been partially characterised by ARDRA, SDS-PAGE and 16S rRNA gene sequencing (Zakhia *et al*. 2004). Genes for MLSA were chosen according to previous taxonomic and phylogenetic studies: *recA* (Recombinase A) and *atpD* (ATP synthase beta subunit) were used in phylogenetic study of *Agrobacterium* and *Rhizobium* (Gaunt *et al*., 2001); *glnA* (Glutamine synthase) was used by Turner & Young (2000) and *gltA* (Citrate synthase) was used by Hernandez-Lucas *et*

al. (2004). In a study evaluating the taxonomic potential of 10 housekeeping genes, Martens *et al.* (2007, 2008) found these genes and additionally *thrC* (threonine synthase) usefull for identification and inference of phylogenetic relationships of *Ensifer* species. Phylogenetic analyses of the five housekeeping genes were compared with the 16SrRNA gene tree using three algorithmic methods (ML, MP and NJ). We completed the characterisation with auxanographic tests and *nodA* gene sequences for biovar determination. The *nodA* gene is a key gene for establishment of symbiosis; it is present as a single copy and has a size of 590 to 660 bp (Moulin *et al.*, 2004). Except for some photosynthetic bradyrhizobia (Giraud *et al.*, 2007) all bacteria described thus far capable of establishing a symbiosis with legumes, harbour nodulation (*nod*) genes and nitrogen fixation (*nif, fix*) genes, opening the prospect of screening these genes as an alternative to nodulation tests. The encoded αβ-unsaturated Nod factor protein is a key enzyme for Nod Factor synthesis, transferring an acyl chain on a chitin oligomer, resulting in a biologically active NodA molecule. *nodA* gene sequences provide information on the symbiotic characteristics of the rhizobium and may be predictive for the type of Nod factor produced and the host specificity (Lortet *et al.*, 1996; Debellé *et al.*, 2001). Because of its frequent plasmid-borne nature and its propension to lateral gene transfer, its sequence is usually indicative of biovar rather than species affiliation (Haukka *et al.*, 1998).
All phylogenies applying any of the three algorithmic methods unambiguously placed the nine strains within the Alphaproteobacteria, in the *Ensifer* clade. The nine strains were subdivided into three monophyletic clusters with high intra-species sequence similarity of 100 % for all gene sequence analyses.

Cluster A formed a separate cluster with variable positions relative to the *Ensifer* reference species in the phylogenies of the 16S rRNA and all housekeeping genes. In the combined gene sequence analysis Cluster A formed a single, seperate cluster which was clearly distinguished from all other clusters. The sequence homologies between Cluster A strains and strains of related reference species supported this segregation. The sequence similarity values between cluster A and the other genotypic clusters ranged between 82-94.3 % in the different gene analyses (Fig. 2). *S. chiapanecum* sp. nov., related to *S. terangae*, was reported during the process of revision of the present work (Rincón-Rosales et *al.*, 2009). We thus retrieved the available recA and 16S rRNA gene sequences of this species and observed that both genes place *S. chiapanecum*, *S. terangae* and *S. mexicanus* in a cluster (96.3-96.9% (*recA*) and 99.3-99.7% (16S RNA gene) internal similarity values), away from cluster A, with interspecies similarity values ranging from 92.2 to 92.6% (recA) and 98.4 to 98.9% (16S RNA).

These values illustrate a clear gap for sequence similarity levels within cluster A strains and between cluster A and other reference strains, corresponding to the inter- and intra-species sequence similarity value gap observed by Martens *et al*. (2007, 2008) in their comparison of MLSA data and DNA-DNA hybridizations. This indicates that cluster A strains are well distinguished from the *Ensifer* reference species and therefore represent a new genomic species.

The auxanographic tests confirmed that Cluster A clearly formed a novel *Ensifer* species. We also analysed the *nodA* sequences of the Cluster A representative strain ORS 1444 to get some information on the Nod factor of *Argyrolobium uniflorum* microsymbionts. The results showed that it grouped in a separate cluster close to *E. meliloti*/*E. medicae nodA* group. In view of our phenotypic and genotypic results, and those obtained in previous work (Zakhia *et al*., 2004), we propose to create a novel *Ensifer* species for Cluster A strains with the name *Ensifer numidicus* sp. nov., with ORS 1407 as the type strain. In a parallel and independent work, Mahdhi *et al*. (2008) described root nodule isolates from *Argyrolobium uniflorum* in the same region of Tunisia. Among the most effective strains (potential candidates for inoculant production), 13 formed a homogeneous and separate group on the *Ensifer* branch by 16S rDNA PCR-RFLP. Three representative strains of this group were sequenced, and they shared identical sequence with strain ORS 1444. This group may thus be considered as *E. numidicus*.

Cluster B formed a monophyletic group in all phylogenies and grouped seperately with *E. adhaerens* gv. A in the majority of the gene phylogenies. In the 16S rRNA gene tree, Cluster B strains were grouped together with the *E. adhaerens* genomovars and "*S. morelense*". In the single and combined housekeeping gene analyses, Cluster B grouped with *Ensifer adhaerens* gv. A, supported by high bootstrap values. The analysis of the sequence similarities placed Cluster B in the gap between intra and interspecies sequence similarities values, confirming that cluster B belongs to *E. adhaerens* species but with uncertain genomovar position. Following the recommendation of Willems *et al*. (2003) for genomovar discrimination, we performed DNA/DNA hybridisations. The results confirmed that Cluster B is different from genomovars B and C but belongs to *E. adhaerens* genomovar A. Phenotypic and genotypic analyses clearly indicate that cluster B groups with *E. adhaerens* gv. A. All strains from Cluster B nodulate their plant of isolation *Lotus arabicus*. This is in contrast to all known members of the *E. adhaerens* species (Casida, 1982; Willems *et al.*, 2003), even though Rogel *et al*. (2001) demonstrated that they may acquire nodulation capacity upon introduction of symbiotic plasmids.

Cluster C formed a separate monophyletic group in all phylogenies, essentially in the *E. terangae* / *E. mexicanum* clade. A similar grouping was obtained using the four congruent genes (*glnA, gltA, thrC* and *recA)*, in individual gene and in the concatenated gene trees. Sequence similarity values indicate a clear gap between similarity levels within cluster C strains and between cluster C and other reference strains. Values between cluster C strains and the reference strains were comparable with those at the interspecies level found by Martens *et al.* (2008) in their study comparing MLSA data and DNA-DNA hybridizations. This indicates that Cluster C is distinct and forms a separate genospecies. Also the numerical analysis of auxanographic results confirmed that cluster C is separate from all *Ensifer* species. Cluster C strains can use malonate, but not *p*-hydroxybenzoate in contrast to its closest neighbour *E. terangae* LMG 7834^{T}. Two representative strains of cluster C, ORS 1400 and STM 354, isolated from two different legumes, exhibited *nodA* sequences which grouped together (99 % bootstrap value) with those from Cluster A and B, in *E. meliloti* and *E. medicae* branch. The combination of the results obtained by Zakhia *et al*. (2004) by SDS-PAGE of whole cell protein and 16S rRNA ARDRA and in this study by MLSA and additional 16S rRNA and *nodA* gene sequence analyses, auxanographic test indicate that cluster C represents a novel genospecies in the *E. terangae, E. mexicanum* clade, for which we propose the the name: *Ensifer garamanticus* sp. nov., with ORS 1400 as the type strain.

We did not perform additional DNA-DNA hybridizations to establish the separate species status of clusters A and C because of the clear MLSA evidence. In this study the highest sequence similarity level with other *Ensifer* species we obtained for cluster A was 94.3% (atpD) and for cluster C this was 96.4% (gltA). In view of the finding that strains of cluster B which had housekeeping gene similarity levels of 96.4 to 98.4% with *E. adhaerens* genomovars B and C, yield DNA-binding levels below the species threshold (Table 2, 47 to 60%), there will be even less DNA-binding between strains that have lower sequence similarity levels in the housekeeping genes. This information is also in line with a previous elaborate comparison of housekeeping gene sequences and DNA-DNA hybridizations (Martens et al., 2008) where it was shown that MLSA of selected housekeeping genes can accurately predict relations between closely related organims. In view of all these findings we conclude from our data that cluster A and C represent two new *Ensifer* species, for which we propose the names of *E. numidicus* and *E. garamanticus* respectively.

Description of *Ensifer numidicus* sp. nov. (Cluster A)

Ensifer numidicus (nu.mi'di.cus. L. masc. adj. numidicus, pertaining to the country of Numidia, Numidian*numidicus*numidicus, the roman denomination of the region in North-West Africa from which the majority of the organisms were isolated).

Short, aerobic, Gram-negative, non-spore-forming rods. Strains grow on Yeast Mannitol medium, on which they form white slightly mucuous colonies after a growth period of three days at 28 °C.

At the molecular level, this species can be differentiated by phylogenetic analysis based on several housekeeping gene (*recA, glnA, gltA, thrC, atpD*) and 16S rRNA gene sequencing. Its phylogenetic neighbors are *E. medicae, E. meliloti* and *E. arboris*. Detailed phenotypic features for all strains are given in Supplementary Table S2. *Ensifer numidicus* can be distinguished from *E. medicae*, *E. meliloti* and *E. arboris* by differential use of a combination of growth substrates. *Ensifer numidicus* strains do not grow on xylitol and DL-glycerate. Results for other substrates were variable between members of this cluster. About 28 substrates were differently used by strains ORS 1444, ORS 1410 and ORS 1407 and 13 substrates were only metabolised by ORS 1407 (dulcitol, D-lyxose, 1-O-methyl-α-D-glucopyranoside, 3-O-Methyl-D-glucopyranose, D-gluconate, L-histidine, succinate, fumarate, ethanolamine, DL-β-hydroxybutyrate, L-aspartate, L-alanine, propionate). Thus *Ensifer numidicus* strains cannot be identified by biochemical and physiological characters alone. Strains can nodulate *Argyrolobium uniflorum* (ORS 1444 and ORS 1407) and *Lotus creticus* (ORS 1410). The nodulation gene *nodA* differs in sequence compared to those of other *Ensifer* species but is most closely related to those of the *E. meliloti* and *E. medicae* clade.

The type strain of this species is ORS 1407^T (LMG 24690^T, CIP 109850^T). The %GC content of its DNA is 62.8 % (HPLC).

A group of 13 *Sinorhizobium* sp. strains reported by Mahdhi *et al.* (2008), AB1, AB3, AB5, AB6, AB10, AM1, AM2, AM3, AS1, AS2, AS3, AS4, AS5, may be considered as *Ensifer numidicus* and included as members of this species. These strains were independantly isolated from *Argyrolobium uniflorum* in South Tunisia. These strains form a homogeneous group by PCR-RFLP of 16S rRNA gene. Three representative strains of this group STM 4034 (AB1), STM 4036 (AB3), STM 4039 (AS1), share identical 16S rRNA gene sequence with *E. numidicus* strains. The three latter strains form the most effective symbiosis with *Argyrlobium uniflorum* with STM 4036 as the most efficient one. They tolerate pH 6-9, 1-2 % NaCl and 40°C for growth. They are sensitive to ampicillin (100 μg ml^{-1}) and

Streptomycin (100 μg ml^{-1}) but resistant to Kanamycin (100 μg ml^{-1}) and nalixidic Acid (100 μg ml^{-1}).

Description of *Ensifer garamanticus* sp. nov. (Cluster C)

Ensifer garamanticus (ga.ra.man.'ti.cus. L. masc. adj. garamanticus, pertaining to Garamante, Garamantian, *garamanticus*, "the country of Garamantes". In Roman times, Garamantes were inhabitants living in the region South Numidia (Tunisia). Garamantis was their eponym hero. Strains were isolated in this region, which is semi-arid and sunny.

Short, aerobic, Gram-negative, non-spore-forming rods. The strains form white and slightly mucous colonies on YMA medium after 48 to 72 h incubation at 28 °C.

Strains efficiently nodulate *Argyrolobium uniflorum* (ORS 1400, ORS 1401) and *Medicago sativa* (STM 354).

Ensifer garamanticus can be distinguished from other species by phylogenetic analysis based on several housekeeping (*recA, glnA, gltA, thrC,* atpD) and 16S rRNA gene sequencing. Malonate is used by all strains for growth but not by its closest phylogenetic neighbor *E. terangae*. Detailed phenotypic features for all strains are given in Supplementary Table S2. *Ensifer garamanticus*, close to *E. terangae* in phylogenetic analysis, differed in ten substrates used for growth by *E. terangae* LMG 7834^T and not by any member of *Ensifer garamanticus*. However, malonate is metabolised by all *Ensifer garamanticus* strains and not by *E. terangae* LMG 7834^T.

The type strain of this species is ORS 1400^T (LMG 24692^T, CIP 109916^T). The %GC content of its DNA is 62.4 % (HPLC).

The group formed by *Sinorhizobium* sp. strains STM 4015, STM 4016, STM 4027, STM 4031, STM 4032, isolated from *Genista saharae* in South Tunisia by Mahdhi *et al.* (2007), share identical 16S rRNA gene sequence with *E. garamanticus* type strain. They may be thus be considered as *Ensifer garamanticus* members. These strains were described also phenotypically and generally tolerate high temperature (40°C), high pHs (7-12) and high NaCl concentrations (1% up to 4%) for growth. They nodulate their plant of isolation, so that *Genista saharae* may be considered as potential host plant of *Ensifer garamanticus*.

Abbreviations

CIP: Institut Pasteur (Paris) Collection

Acknowledgements

The authors are grateful to F. Boukhatem, B. Dreyfus, A., A. Galiana, L. Moulin, G. Béna, B. Hamed Bey, H. Ameziane and L. Bouchentouf for helpful discussions. C. M., F. Z., A. B. and A. S. are indebted to the *Département Soutien-Formation* (DSF) of the *Institut de Recherche pour le Développement* (IRD), France for PhD fellowship and ESCD grants and M. Mosbah for AUF grant. AW is grateful to the Fund for Scientific Research – Flanders for a postdoctoral fellowship. This work was supported by the Commission of the European Communities RT-program contract BACDIVERS QLRT-2001-02097.

REFERENCES

Baele, M., Baele, P., Vaneechoutte, M., Storms, V., Butaye, P., Devriese, L. A., Verschraegen, G., Gillis, M. & Haesebrouck, F. (2000). Application of tRNA intergenic spacer PCR for identification of *Enterococcus* species. *J Clin Microbiol* **38**, 4201-4207.

Bailly, X., Olivieri, I., de Mita, S., Cleyet-Marel, J-C. & Béna, G. (2006). Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*. *Mol Ecol* **15**, 2719-34.

Bertrand, H. (1997). Stimulation de l'absorption minérale et de la croissance du Colza inoculé avec des bactéries isolées de son rhizoplan. Thèse de l'Université Claude Bernard-Lyon I.

Casida, L. E., Jr. (1982). *Ensifer adhaerens*, gen. nov., sp. nov. a bacterial predator of bacteria in soil. *Int J Syst Bacteriol* **32**, 339–345.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P. & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147-167.

Debellé, F., Moulin, L., Mangin, B., Dénarié, J. & Boivin, C. (2001). *nod* genes and Nod signals and the evolution of the rhizobium legume symbiosis. *Acta Biochem Pol* **48**, 359-365.

de Lajudie, P., A. Willems, B. Pot, D. Dewettinck, G. Maestrojuan, M. Neyra, M. D. Collins, B. L. Dreyfus, K. Kersters & Gillis, M. (1994). Polyphasic taxonomy of rhizobia. Emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov. and *Sinorhizobium teranga* sp. nov. *Int. J. Syst. Bacteriol*. **44,** 715-733.

Eardly, B. D., Nour, S. M., van Berkum, P. & Selander, R. K. (2005). Rhizobial 16S rRNA and *dnaK* genes: Mosaicim and the uncertain phylogenetic placement of Rhizobium galegae. *Appl Env Microbiol* **27**,1328- 1335.

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224-229.

Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1995). Testing significance of incongruence. *Cladistics* **10**, 315-319.

Ferchichi (1996). La lutte contre l'ensablement et pour la stabilisation des dunes : essai de la fixation biologique des dunes en Tunisie presaharienne. *Recherche sur la désertification dans la Jeffara. Rev. Tunis. Geogr.* ***12,*** *49-102.*

Gaunt, M. W, Turner, S. L, Rigottier-Gois, L., Lloyd-Macgilp, S. A. & Young, J. P. (2001). Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Microbiol* **51**, 2037-2048.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L. & Swings, J. (2005). Opinion: Re-evaluating prokaryotic species. *Nature Rev Microbiol* **3**, 733-739.

Giraud E., Moulin L., Vallenet D., & 31 other authors (2007). Legume symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. *Science* **316**, 1296-1297.

Goris, J., Suzuki, K-I., De Vos, P., Nakase, T. & Kersters, K. (1998). Evaluation of microplate DNA-DNA hybridization method compared with the initial renaturation method. *Can J Microbiol* **44**, 1148-1153.

Haukka, K., Lindström, K. & Young, J. P. (1998). Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. *Appl Environ Microbiol* **64**, 419-26.

Hernandez-Lucas, I., Rogel-Hernandez, M. A., Segovia, L., Rojas-Jiménez, K. & Martinez-Romero, E. (2004). Phylogenetic relationships of rhizobia based on citrate synthase gene sequences. *Syst Appl Microbiol* **27,** 703-706.

Judicial Commission of the International Committee on Systematics of Prokaryotes (2008). The genus name *Sinorhizobium* Chen *et al*. 1988 is a later synonym of *Ensifer* Casida 1982 and is not conserved over the latter genus name, and the species name '*Sinorhizobium adhaerens*' is not validly published. Opinion 84. *Int J Syst Evol Microbiol* **58**, 1973.

Kersters, K. & De Ley, J. (1984). Genus III. *Agrobacterium* Conn 1942. In *Bergey's manual of systematic bacteriology*, vol. 1 pp. 244-254. Edited by N. R. Krieg & J. G. Holt. The Williams & Wilkins Co., Baltimore.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120.

Lloret, L., Ormeno-Orrillo, E., Rincon, R., Martinez-Romero, J., Rogel-Hernandez, M. A. & Martinez-Romero, E. (2007). *Ensifer mexicanus* sp. nov. a new species nodulating *Acacia angustissima* (Mill.) Kuntze in Mexico. *Syst Appl Microbiol* **30**, 280-290.

Lortet, G., Mear, N., Lorquin, J., Dreyfus, B., de Lajudie, P., Rosenberg, C. & Boivin, C. (1996). Nod factor thin-layer chromatography profiling as a tool to characterize symbiotic specificity of rhizobial strains : application to *Sinorhizobium saheli, S. teranga* and *Rhizobium* sp. strains isolated from *Acacia* and *Sesbania*. *Mol. Plant-Microbe Interact* **9:**736-747.

Mahdhi, M., de Lajudie, P. & Mars, M. (2008). Phylogenetic and symbiotic characterization of rhizobial bacteria nodulating *Argylorobium uniflorum* in Tunisian arid soils. *Can J Microbiol* **54**, 209-217.

Mahdhi, M., Nzoué, A., Gueye, F., Merabet, C., de Lajudie, P. & Mars, M. (2007). Phenotypic and genotypic diversity of *Genista saharae* microsymbionts from the infra-arid region of Tunisia. *Lett Appl Microbiol* ISSN 0266-8254.

Ma**ntelin, S., Fischer-Le Saux, M., Zakhia, F., Béna, G., Jeder, H., de Lajudie, P. & Cleyet-Marel J. C. (2006).** Emended description of the genus *Phyllobacterium* and description of four novel species associated to plant roots

Phyllobacterium bourgognense sp. nov. *Phyllobacterium ifriqiyense* sp. nov. *Phyllobacterium leguminum* sp. nov. *and Phyllobacterium brassicacearum* sp. nov. *Int J Syst Evol Microbiol* **56**, 827-839.

Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57**, 489-503.

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. & Willems, A. (2008). Advantages of MultiLocus Sequence Analysis for taxonomic studies: A case study using 10 housekeeping genes in the genus *Ensifer*. *Int J Syst Evol Microbiol* **58**, 200-214.

Moulin, L., Béna, G., Boivin-Masson, C. & Stepkowski T. (2004). Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the Bradyrhizobium genus. *Mol Phyl Evol* 30(3):720-32.

Naser, S. M., Thompson, F. L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M. & Swings, J. (2005). Application of Multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiol*. **151**, 2141-2150.

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., Evtushendro, L. & Misra, A. K. (1996). Molecular phylogeny of the genus *Frankia* and related genera and emendation of the Family Frankiaceae. *Int J Syst Bact* **46**, 1-9.

Normand, P., Cournoyer, B., Simonet, P. & Nazaret, S. (1992). Analysis of a ribosomal RNA operon in the actinomycete *Frankia*. *Gene* **111**, 119-124.

Posada, D. & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of the Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**, 793–808.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Rincon-Rosales, R., Lloret, L., Ponce, E. & Martínez-Romero, E. (2009). Rhizobia with different symbiotic efficiencies nodulate *Acaciella angustissima* in Mexico, including *Sinorhizobium chiapanecum* sp. nov. which has common symbiotic genes with *Sinorhizobium mexicanum*. *FEMS Microbiol. Ecol*. **67,** 103 – 117.

Rogel, M. A., Hernandez-Lucas, I., Kuykendall, L. D., Balkwill, D. L. & Martinez-Romero, E. (2001). Nitrogen-fixing nodules with *Ensifer adhaerens* harbouring *Rhizobium tropici* symbiotic plasmids. *Appl Environ Microbiol* **67**, 3264-3268.

Saitou, N. & Nei, M. (1987). The neighbour joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

Schouls, L. M., Schot, C. S. & Jacobs, J. A. (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *J Bacteriol* **185**, 7241-7246.

Stackebrandt, E. (2003). The richness of prokaryotic diversity: there must be a species somewhere. *Food Technol Biotechnol* **41**, 17-22.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. Kampfer, P., Maiden, M. C., Nesme, X., Rossello-Mora, R., Swings, J., Trupper, H. G., Vauterin, L., Ward, A. C. & Whitman, W. B. (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043-1047.

Stepkowski, T., Czaplinska, M., Miedzinska, K. & Moulin, L. (2003): The variable part of the *dnaK* gene as an alternative marker for phylogenetic studies of rhizobia and related alpha Proteobacteria. Syst. Appl. Microbiol. 26, 483-494.

Soler, L., Yanez, M. A., Chacon, M. R., Aguilera-Arreola, M. G., Catalan, V., Figueras, M. J. & Martinez-Murcia, A. J. (2004). Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *Int J Syst Bacteriol* **54**, 1511-1519.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignement aided by quality analysis tools. *Nucleic Acids Res* **24**, 4876-4882.

Toledo, I., LIoret, L. & Martinez-Romero, E. (2003). *Sinorhizobium americanum* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. *Syst Appl Microbiol* **26**, 54-64.

Turner, S. L. & Young, J. P. W. (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* **17**, 309-319.

Ueda, K., Seki, T., Kudo, T., Yoshida, T. & Kataoka, M. (1999). Two distinct mechanisms cause heterogeneity of 16S rRNA. *J Bacteriol* **181**, 78-82.

van Berkum, P., Terefework, Z., Paulin, L., Suomalainen, S. Lindström, K. & Eardly, B. D. (2003). Discordant phylogenies within the rrn loci of rhizobia. *J Bacteriol* 185, 2988-2998.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407-438.

Vinuesa, P., Silva, C., Lorite, M. J., Izaguirre-Mayoral, M. L., J.Bedmar, E. & Martinez-Romero, E. (2005a). Molecular systematics of rhizobia based on maximum likelihood and Baysian phylogenies inferres from rrs, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. *Syst Appl Micobiol* **28**, 702-716.

Vinuesa, P., Silva, C., Werner, D. & Martinez-Romero, E. (2005b). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Mol Phylogenet Evol* **34**, 29-54.

Vinuesa, P., Leon-Barrios, M., Silva, C., Willems, A., Jarabo-Lorenzo, A., Pérez-Galdona, R. & Martinez-Romero, E. (2005c). *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (*Papilionoideae :Genisteae*) growing in the canary Islands, along with *B. japonicum* bv. *genistearum*, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β. *Int J Syst Evol Microbiol* **55**, 569-575.

Wang, E. T., Tan, Z. Y., Willems, A., Fernandez-Lopez, M., Reinhold-Hurek, B. & Martinez-Romero, E. (2002). *Sinorhizobium morelense* sp. nov., isolated from *Leucaena leucocephala* that is highly resistant to multiple antibiotics. *Int J Syst Microbiol* **52**, 1687-1693.

Wei, G. H., Wang, E. T., Tan, Z. Y., Zhu, M. E. & Chen, W. X. (2002). *Rhizobium indigoferae* sp. nov. and *Sinorhizobium kummerowiae* sp. nov., respectively isolated from *Indigofera* spp ; and *Kummerowia stipulacea*. *Int J Syst Evol Microbiol* **52**, 2231-2239.

Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. (2003). A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* **16**, 1236-1248.

Willems, A., Fernandez-Lopez, M., Munoz-Adelantado, E., Goris, J., De Vos, P., Martinez-Romero, E., Toro, N. & Gillis, M. (2003). Description of new *Ensifer* strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an Opinion. *Int J Syst Evol Microbiol* **53**, 1207-1217.

Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., De Vos, P. & Gillis, M. (2001). DNA-DNA hybridization study of *Bradyrhizobium* strains. *Int J Syst Microbiol* **51**, 1315-1322.

Young, J. M. (2003). The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen *et al*. 1988, and *Sinorhizobium morelense* Wang *et al*. 2002 is a later synonym of *Ensifer* Casida 1982. Is the combination *Sinorhizobium adhaerens* (Casida 1982) Willems *et al*., 2003 legitimate? Request for an Opinion. *Int J Syst Evol Microbiol* **53**, 2107-2110.

Zakhia, F., Jeder, H., Domergue, O., Willems, A., Cleyet-Marel, J.-C. , Gillis, M., Dreyfus, B. & de Lajudie, P. (2004). Characterisation of Legume-Nodulating Bacteria (LNB) in arid regions of Tunisia. *Syst Appl Microbiol* **27**, 380-395.

Zakhia, F. (2004). Diversité des bactéries hôtes de légumineuses méditerranéennes en Tunisie et au Liban. Thèse d'université. Université de Montpellier, France. 200 p.

Zakhia, F., Jeder', H., Willems, A., Dreyfus, B. & de Lajudie, P. (2006). Diverse bacteria associated with root nodules of spontaneous legumes in Tunisia and first report for *nifH*-like gene within the genera *Microbacterium* and *Starkeya*.. *Microbiol. Ecol*. **51**(3), 375-393.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* **53**, 1893-1900.

Table 1. Accession numbers of new sequence data

ND, not detected.

Name	Strain	Other number	SSU	*recA*	*glnA*	*atpD*	*thrC*	*gltA*	*nodA*	Source
E. numidicus	ORS 1444	LMG 24691	AY500253	AM946578	AM946562	AM946549	AM946586	AM946596		*Argyrolobium uniflorum,* Tunisia (Zakhia *et al*., 2004)
E. numidicus	ORS 1410	LMG 24736, CIP 109858	AM946566	AM946577	AM946561	AM946550	ND	AM946595		*Lotus creticus,* Tunisia (Zakhia *et al.*, 2004)
E. numidicus	ORS 1407^{T}	LMG24690^{T}, CIP 109850^{T}	AY500254	AM946576	AM946560	AM946551	AM946585	AM946594		*Argyrolobium uniflorum,* Tunisia (Zakhia *et al*., 2004)
E. garamanticus	ORS 1400^{T}	LMG 24692^{T}	AY500255	AM946573	AM946557	AM946546	AM946582	AM946591		*Argyrolobium uniflorum,* Tunisia (Zakhia *et al*., 2004)
E. garamanticus	ORS 1401	LMG 24693, CIP 109848	AM946567	AM946574	AM946558	AM946547	AM946583	AM946592		*Argyrolobium uniflorum,* Tunisia (Zakhia *et al*., 2004)
E. garamanticus	STM 354	LMG 24694, CIP 109849	AM946568	AM946575	AM946559	AM946548	AM946584	AM946593		*Medicago sativa,* Tunisia (Zakhia *et al.*, 2004)
E. adhaerens gv. A	ORS 2154		AM946569	AM946579	AM946565	AM946553	AM946587	AM946598		*Lotus arabicus,* Senegal (Sy *et al.*, 2001)
E. adhaerens gv. A	ORS 2133		AM946570	AM946580	AM946563	AM946554	AM946588	AM946597		*Lotus arabicus,* Senegal (Sy *et al.*, 2001)
E. adhaerens gv. A	ORS 529		AM946571	AM946581	AM946564	AM946555	AM946589	AM946599		*Lotus arabicus,* Senegal (Sy *et al.*, 2001)

Table 2. DNA-DNA hybridisation values among *Ensifer* species and strains of cluster B isolated from *Lotus arabicus* in Senegal.

	ORS 529	**ORS 2133**
ORS 529	**100**	**99**
ORS 2133	**99**	**100**
***Ensifer adhaerens* gv. A LMG 10007**	**74**	**75**
***Ensifer adhaerens* gv. B R-7457**	60	59
***Ensifer adhaerens* gv. C LMG 20216^{T}**	47	47

Supplementary Table S1. Primers used for DNA amplification and sequencing.

Primers given in bold were used for initial amplification.

Gene	Primer	Direction	length (bases)	Primer sequence (5'-3')	Reference
16S rRNA	**FGPS 6**	forward	22	GGA GAG TTA GAT CTT GGC TCA G	Normand *et al.*, 1992
16S rRNA	**FGPS 1509**	reverse	20	AAG GAG GGG ATC CAG CCG CA	Normand *et al.*, 1992
16S rRNA	FGPS 484-292	forward	15	CAG CAG CCG CGG TAA	Normand *et al.*, 1996
16S rRNA	16S-1080r	reverse	19	GGG ACT TAA CCC AAC ATC T	Sy *et al.*, 2001
16S rRNA	16S-870f	forward	21	CCT GGG GAG TAC GGT CGC AAG	Sy *et al.*, 2001
16S rRNA	16S-1924	reverse	20	GGC ACG AAG TTA GCC GGG GC	Sy *et al.*, 2001
recA	**A555**	reverse	26	CGR ATC TGG TTG ATG AAG ATC ACC AT	Gaunt *et al.*, 2001
recA	**A36**	forward	23	ATC GAG CGG TCG TTC GGC AAG GG	Gaunt *et al.*, 2001
atpD	**atpD273F**	forward	21	SCT GGG SCG YAT CMT GAA CGT	Gaunt *et al.*, 2001
atpD	**atpD771R**	reverse	23	GCC GAC ACT TCC GAA CCN GCC TG	Gaunt *et al.*, 2001
glnA	**glnA144F**	forward	21	GTC ATG TTC GAC GGY TCY TCG	Martens *et al*., 2007
glnA	**glnA1142R**	reverse	21	TGG AKC TTG TTC TTG ATG CCG	Martens *et al*., 2007
glnA	glnA572F	forward	20	GGA CAT GCG YTC YGA RAT GC	Martens *et al*., 2007
glnA	glnA572R	reverse	20	GCA TYT CRG ARC GCA TGT CC	Martens *et al*., 2007
thrC	**thrC577F**	forward	21	GGC AMK TTC GAC GAY TGC CAG	Martens *et al*., 2007
thrC	**thrC1231R**	reverse	20	GGR AAT TTD GCC GGR TGS GC	Martens *et al*., 2007
thrC	thrC766F	forward	17	GGC AAT TTC GGC GAY AT	Martens *et al*., 2007
thrC	thrC766R	reverse	17	ATR TCG CCG AAA TTG CC	Martens *et al*., 2007
thrC	thrC925R	reverse	20	GAS GAR AYC TGG ATR TCC AT	Martens *et al*., 2007
gltA	**gltA428F**	forward	19	CSG CCT TCT AYC AYG ACT C	Martens *et al*., 2007
gltA	**gltA1111R**	reverse	20	GGG AAG CCS AKC GCC TTC AG	Martens *et al*., 2007
nodA	NodA1F	forward	24	TGC RGT GGA ARN TRB VYT GGG AAA	Haukka *et al.*, 1998
nodA	NodAB1R	reverse	23	GGN CCG TCR TCR AAW GTC ARG TA	Haukka *et al.*, 1998

Supplementary Table S2: Carbon assimilation tests. Cluster A is compared to reference strains of *E. meliloti*, *E. medicae* and *E. arboris* ; Cluster B is compared to *E. adhaerens* genomovars A, B, C ; Cluster C is compared to *E. terangae* reference strain. The following substrate of the biotype 100 strip (bioMérieux) are assimilated by all strains: D(+) Galactose, β-D (+) Fructose , D(+) Trehalose, D(+) Mannose, Sucrose (Saccharose), Maltose, α-Lactose, Lactulose , 1-0-Methyl-β-galactopyranoside, 1-0-Methyl-α-galactopyranoside, D(+)Xylose, Palatinose, α-L-Rhamnose, D(+) Arabitol, Glycerol, myo-Inositol, D-Mannitol, Maltitol, D(+)Turanose , D- Sorbitol, L-Glutamate, L-Proline, Betain, .

The following are not assimilated by any strain: D-Saccharate, Mucate, L(+) Tartrate, D(-) Tartrate, meso-Tartrate, Tricarballylate, Gentisate, m-Hydroxybenzoate , 3-Phenylpropionate, m-Coumarate, Histamine, Caprylate, Tryptamine

+: Growth ; +/-: Weak Growth ; - No growth

*: Positive reaction is estimated by a colorated reaction (Esculin “dark brown/ Black”, Hydroxyquinoline β-glucuronide “black point”, L- Tryptophan “Dark orange”, L-Histidine “pink coloration »).

b: Black coloration; j: Yellow coloration; i: not colored; c: colored

Substrate	*E. meliloti* LMG 6133T	*E. medicae* LMG 19920T	*E. arboris* LMG 1419T	Cluster A: ORS 1444	Cluster A: ORS 1410	Cluster A: ORS 1407	*E. adhaerens* gvA LMG 9954	*E. adhaerens* gvB R 7457	*E.adhaerens* gvC LMG 20216	Cluster B: ORS 2133	Cluster B ORS 2154	Cluster B: ORS 529	*E. terangae* LMG 7834T	*Cluster C: ORS 1400*	*Cluster C: ORS 1401*	*Cluster C: STM 354*
control	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
α-D (+) Glucose	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+
L(+) Sorbose	+	+/-	-	-	+	+	-	-	+/-	-	-	-	-	-	-	-
α-D Melibiose	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+/-
D(+) Raffinose	+	-	+	-	+	+	+	+	+	+	+	+	+/-	+	+	+
Maltotriose	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+/-
D(+) Cellobiose	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
β-Gentiobiose	+	+/-	-	-	-	-	-	+	+/-	+/-	+/-	+/-	-	-	+/-	+/-
1-0-Methyl-β-D-glucopyranoside	+	+	+	+	-	+/-	+/-	+	+	+	+	+	+	+	+	+
Esculin*	-b	-b	-b	-b	-b	-b	-b	-b	- b	- b	- b	- b	- b	- b	- b	- b
D(-)Ribose	+	+	+	+	-	+/-	+	+	+/-	+	+	+	+	+/-	+	-
L(+) Arabinose	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+/-	-
α-L(-)Fucose	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+/-
D(+) Melezitose	+	-	+	+	+	+	+	+	+	+	+	+	+/-	-	-	-
L(-) Arabitol	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
Xylitol	+	+	+	-	-	-	+	+	+	+	+	+	+	-	-	-
Dulcitol	+	+	+	-	-	+/-	+/-	-	+/-	-	-	-	-	-	-	-
D-Tagatose	+	+	+	-	+	+	+	-	+	-	-	-	-	-	-	-
Adonitol	+	+/-	+	+	+	-i	+	+	+	+	+	+	+	+	+	+
Hydroxyquinoline-β-glucuronide*	+	-	-i	-i	-	-	-i	-	-i	-	-	-	-	-	-	-i
D-Lyxose	-i	+	+/-	-	-	+	+	+	+	+	+	+	-	-	-	-
i-Erythritol	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
1-O-Methyl-α-D-glucopyranoside	+	-	+	-	-	+	-	+	+	+/-	+/-	+/-	+/-	-	-	-
3-O-Methyl-D-Glucopyranose	+	+/-	+	-	-	+	+	-	-	-	-	-	-	-	-	-
D(+) Malate	-	-	+	-	-	-	+	+/-	+/-	+	+	+/-	+/-	-	-	-
L(-) Malate	+	+	+	+/-	-	+/-	+	+	+	+/-	+	+/-	+	+/-	+/-	-
cis-Aconitate	-	-	+	-	-	-	+	+	+	-	+	+/-	+/-	-	-	-
trans-Aconitate	-	-	-	-	-	-	+	+	-	+	+	+/-	+	+/-	+/-	-
Citrate	-	-	+	-	-	-	-	-	+	-	-	-	+/-	-	-	-
D-Glucoronate	-	-	-	-	-	-	+	+	+	+/-	+	+/-	+/-	-	-	-
D-Galacturonate	-	-	+	-	-	-	+	+	+	+	+	+	+	-	-	-
2-Keto-D-Gluconate	+	+	+/-	+	+	+	-	-	-	-	-	-	+	+	+/-	+
5-Keto-D-Gluconate	-	-	+/-	-	-	-	+	-	-	-	-	-	-	-	-	-
L-Tryptophan*	-i	-	-i	-c	-	-c	+/-c	+c	+/-	+/-c	+c	+c	-i	-i	-	-
N-Acetyl-D-Glucosamine	+	+	+	+/-	+	+/-	+	+	+	+	+	+	+	+	+/-	+/-
D- Gluconate	-	-	+	-	-	+/-	+	-	+	-	-	-	+/-	-	-	-
Phenylacetate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Protocatechuate	+	-j	+	+/-	-	+	+	+	+	+	+	+	+	+	+	+
p-Hydroxybenzoate	-	-	+	-	-	-	-	+	-	+	+	+	+	-	-	-
(-) Quinate	-	+	+	-	+	+	+	+	+	-	-	+	+	+	+	+
Trigonelline	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Putrescine	-	-	-	-	-	-	-	-	-	+	+/-	+	-	-	-	-
DL-α-Amino-n-Butyrate	+	+	+	+/-	-	+	+	+	+	+	+	+	+	+	+	+
DL-Lactate	+	-	+	+	+	+	+	+	+	+	+/-	+	+/-	+	+/-	-
Caprate	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-
L.Histidine*	+	+	+	-	-	+/-	-	+	+	+i	+i	+i	+/-i	+i	+	+/-
Succinate	+	+	+	-	-	+/-	+i	+	+/-	+	+	+	-	-	-	+/-
Fumarate	+	+	+	-	-	+/-	+	+	+/-	+	+	+	+/-	+/-	+/-	+/-
Glutarate	-	-	+	-	-	-	+/-	+	-	-	-	-	-	-	-	-
DL-Glycerate	+/-	+/-	+	-	-	-	+/-	-	+/-	+/-	+	+/-	-	-	+/-	-
DL-α-Amino-n-Valerate	+/-	-	+	-	-	-	+/-	-	+/-	+/-	+	+	-	+/-	+/-	-
Ethanolamine	-	+/-	+	-	-	+/-	+/-	-	+	+/-	+	+/-	-	-	-	-
D-Glucosamine	+	+	+	+	-	+/-	+	+	+/-	+	+	+	+	+	+/-	-
Itaconate	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DL-β-Hydroxybutyrate	-	+	+	-	-	+/-	+	+	+	+	+	+	+	+	+	+
L-Aspartate	+	+	+	-	-	+/-	+	+	+	+	+	+	+/-	+/-	-	-
D-Alanine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L-Alanine	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+/-
L-Serine	-	-	+	-	-	-	+	+	+	+	+	+	+/-	+	+/-	+/-
Malonate	-	+/-	+	+/-	-	+/-	-	-	-	-	-	-	-	+/-	+/-	+/-
Propionate	+	-	+	-	-	+/-	+	+	+/-	+	+	+/-	-	-	-	-
L-Tyrosine	-	-	-	-	-	-	-	-	+/-	-	+/-	+/-	-	-	-	+/-
α-Ketoglutarate	-	-	+/-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 1. Phylogenetic reconstruction based on individual analysis of the 16S rRNA gene. Analysis were conducted using the ML method. BT values of 75 or more (using 100 replicates) are indicated at branching points. Bars, 0.1% estimated substitutions.

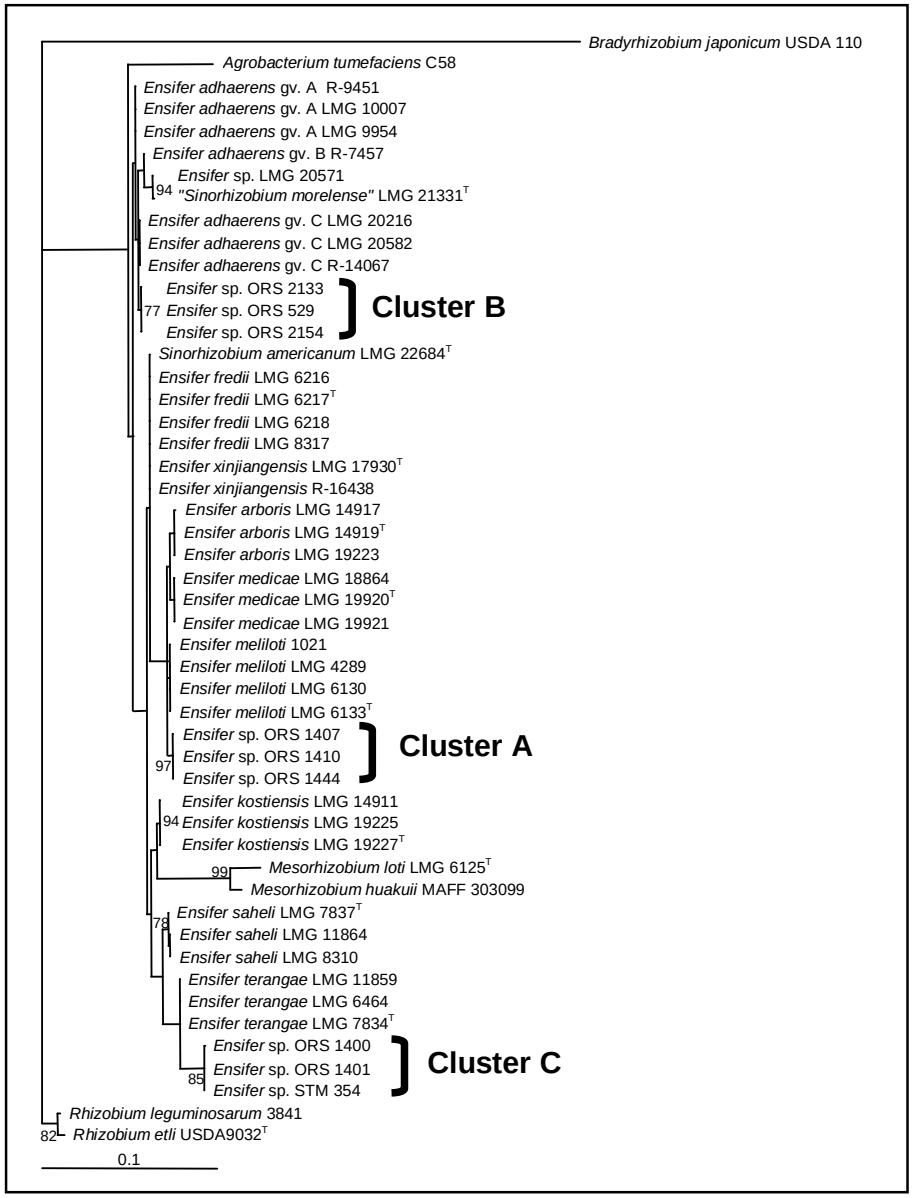
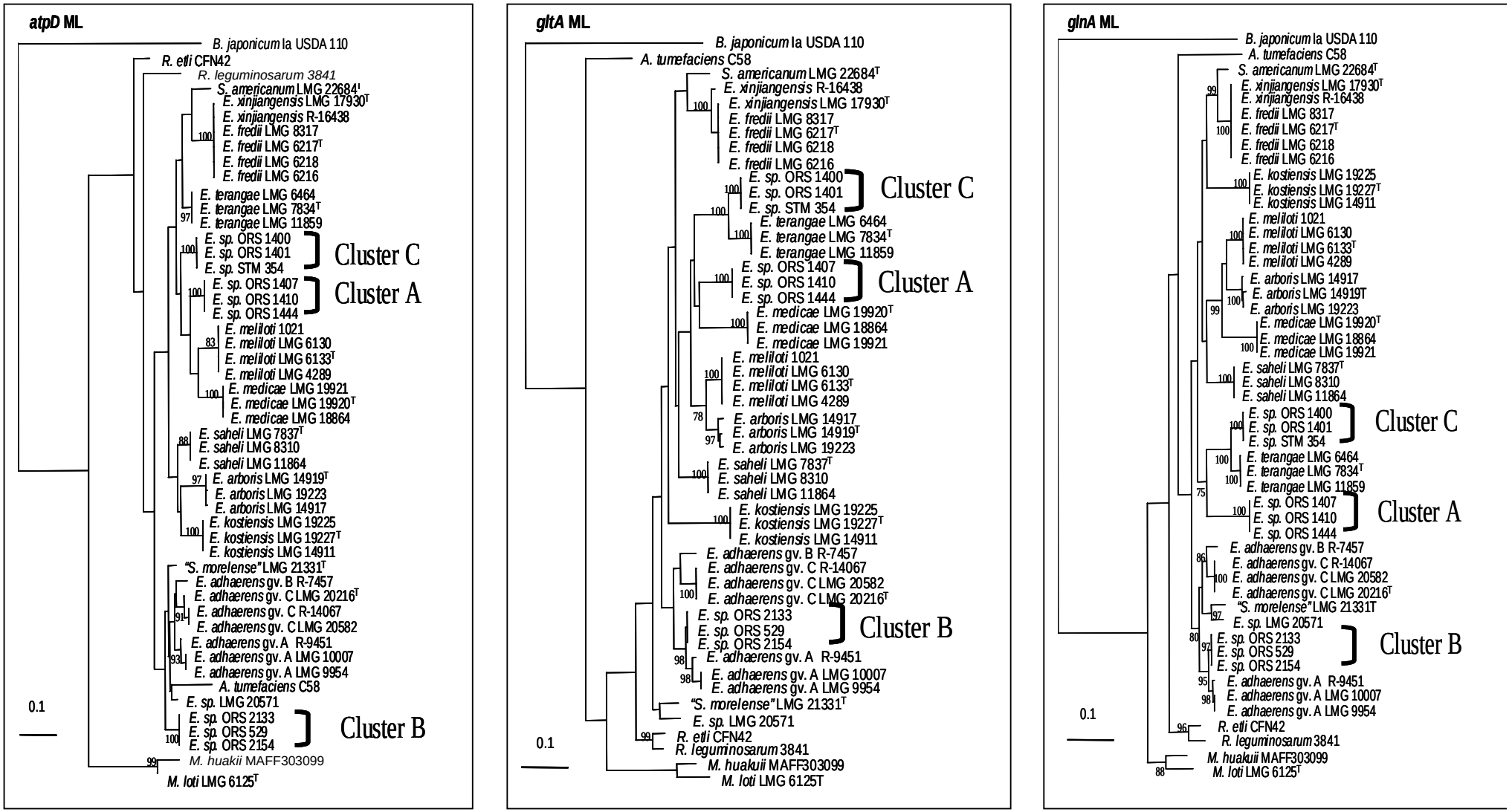
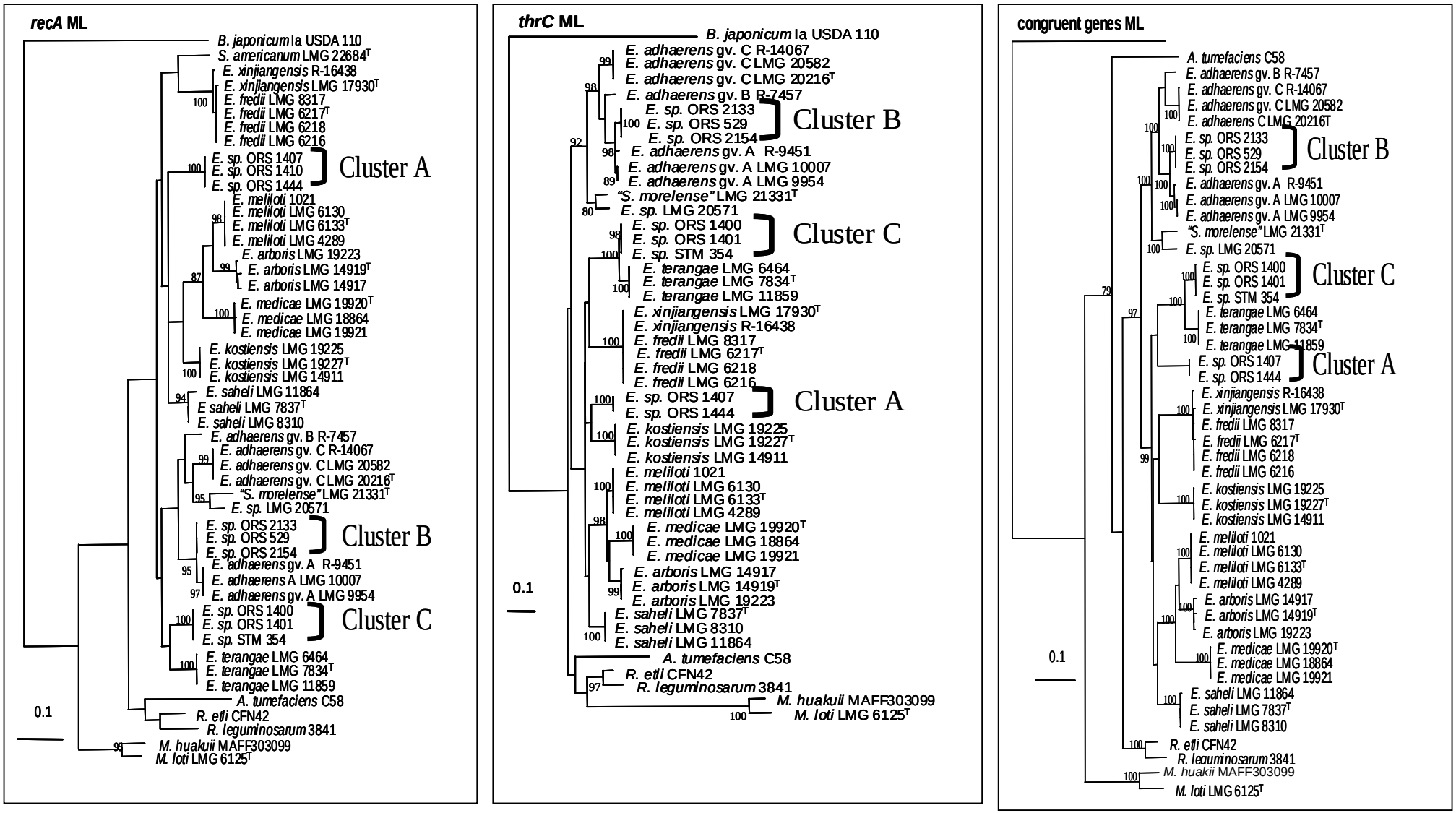


Figure 2: Phylogenetic trees for five housekeeping genes (*atpD, gltA, glnA, recA, thrcC*) and the concatenated tree of Tunisian and Senegalese strains compared with *Ensifer* species reference strains. Trees were calculated using the maximum likelihood method (ML). Bootstrap values of 75 or more derived from 100 replicates are indicated in branching points. Bars, 0.1 % estimated substitutions




recA ML
B. japonicum Ia USDA 110
S. americanum LMG 22684T
E. xinjiangensis R-16438
E. xinjiangensis LMG 17930T
E. fredii LMG 8317
E. fredii LMG 6217T
E. fredii LMG 6218
E. fredii LMG 6216
E. sp. ORS 1407
E. sp. ORS 1410
E. sp. ORS 1444
Cluster A
E. meliloti 1021
E. meliloti LMG 6130
E. meliloti LMG 6133T
E. meliloti LMG 4289
E. arboris LMG 19223
E. arboris LMG 14919T
E. arboris LMG 14917
E. medicae LMG 19920T
E. medicae LMG 18864
E. medicae LMG 19921
E. kostiensis LMG 19225
E. kostiensis LMG 19227T
E. kostiensis LMG 14911
E. saheli LMG 11864
E saheli LMG 7837T
E. saheli LMG 8310
E. adhaerens gv. B R-7457
E. adhaerens gv. C R-14067
E. adhaerens gv. C LMG 20582
E. adhaerens gv. C LMG 20216T
"S. morelense" LMG 21331T
E. sp. LMG 20571
E. sp. ORS 2133
E. sp. ORS 529
E. sp. ORS 2154
Cluster B
E. adhaerens gv. A R-9451
E. adhaerens A LMG 10007
E. adhaerens gv. A LMG 9954
E. sp. ORS 1400
E. sp. ORS 1401
E. sp. STM 354
Cluster C
E. terangae LMG 6464
E. terangae LMG 7834T
E. terangae LMG 11859
A. tumefaciens C58
R. etli CFN42
R. leguminosarum 3841
M. huakuii MAFF303099
M. loti LMG 6125T
0.1
thrC ML
B. japonicum Ia USDA 110
E. adhaerens gv. C R-14067
E. adhaerens gv. C LMG 20582
E. adhaerens gv. C LMG 20216T
E. adhaerens gv. B R-7457
E. sp. ORS 2133
E. sp. ORS 529
E. sp. ORS 2154
Cluster B
E. adhaerens gv. A R-9451
E. adhaerens gv. A LMG 10007
E. adhaerens gv. A LMG 9954
"S. morelense" LMG 21331T
E. sp. LMG 20571
E. sp. ORS 1400
E. sp. ORS 1401
E. sp. STM 354
Cluster C
E. terangae LMG 6464
E. terangae LMG 7834T
E. terangae LMG 11859
E. xinjiangensis LMG 17930T
E. xinjiangensis R-16438
E. fredii LMG 8317
E. fredii LMG 6217T
E. fredii LMG 6218
E. fredii LMG 6216
E. sp. ORS 1407
E. sp. ORS 1444
Cluster A
E. kostiensis LMG 19225
E. kostiensis LMG 19227T
E. kostiensis LMG 14911
E. meliloti 1021
E. meliloti LMG 6130
E. meliloti LMG 6133T
E. meliloti LMG 4289
E. medicae LMG 19920T
E. medicae LMG 18864
E. medicae LMG 19921
E. arboris LMG 14917
E. arboris LMG 14919T
E. arboris LMG 19223
E. saheli LMG 7837T
E. saheli LMG 8310
E. saheli LMG 11864
A. tumefaciens C58
R. etli CFN42
R. leguminosarum 3841
M. huakuii MAFF303099
M. loti LMG 6125T
0.1
congruent genes ML
A. tumefaciens C58
E. adhaerens gv. B R-7457
E. adhaerens gv. C R-14067
E. adhaerens gv. C LMG 20582
E. adhaerens C LMG 20216T
E. sp. ORS 2133
E. sp. ORS 529
E. sp. ORS 2154
Cluster B
E. adhaerens gv. A R-9451
E. adhaerens gv. A LMG 10007
E. adhaerens gv. A LMG 9954
"S. morelense" LMG 21331T
E. sp. LMG 20571
E. sp. ORS 1400
E. sp. ORS 1401
E. sp. STM 354
Cluster C
E. terangae LMG 6464
E. terangae LMG 7834T
E. terangae LMG 11859
E. sp. ORS 1407
E. sp. ORS 1444
Cluster A
E. xinjiangensis R-16438
E. xinjiangensis LMG 17930T
E. fredii LMG 8317
E. fredii LMG 6217T
E. fredii LMG 6218
E. fredii LMG 6216
E. kostiensis LMG 19225
E. kostiensis LMG 19227T
E. kostiensis LMG 14911
E. meliloti 1021
E. meliloti LMG 6130
E. meliloti LMG 6133T
E. meliloti LMG 4289
E. arboris LMG 14917
E. arboris LMG 14919T
E. arboris LMG 19223
E. medicae LMG 19920T
E. medicae LMG 18864
E. medicae LMG 19921
E. saheli LMG 11864
E. saheli LMG 7837T
E. saheli LMG 8310
R. etli CFN42
R. leguminosarum 3841
M. huakii MAFF303099
M. loti LMG 6125T
0.1

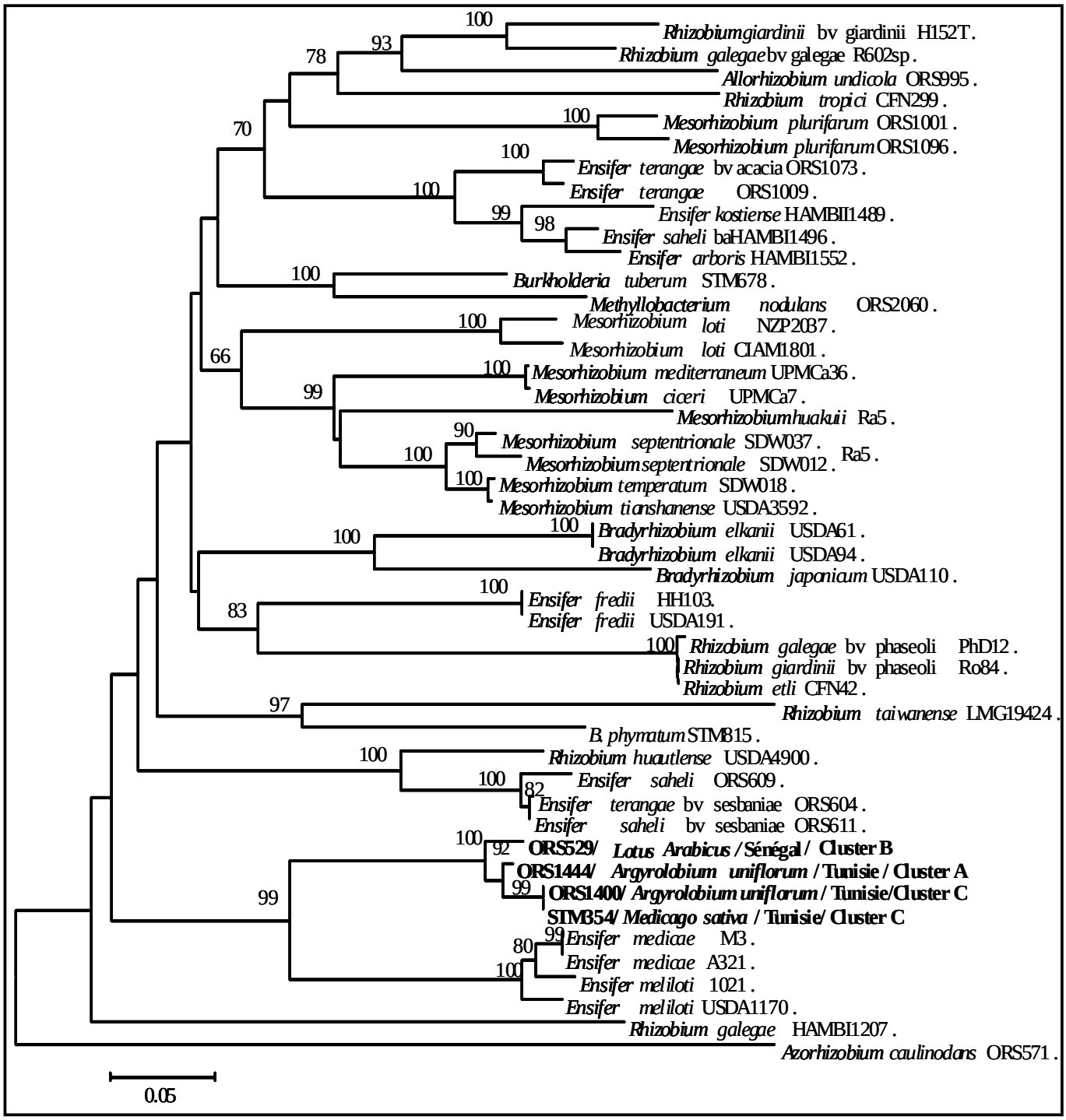


Figure 3: Neighbor-joining phylogenetic tree of *nodA* gene sequences of Tunisian end Senegalease strains. The new strains are indicated in blot and bootstrap value> 60 % resulting from 1000 replicates are indicated in branching points.